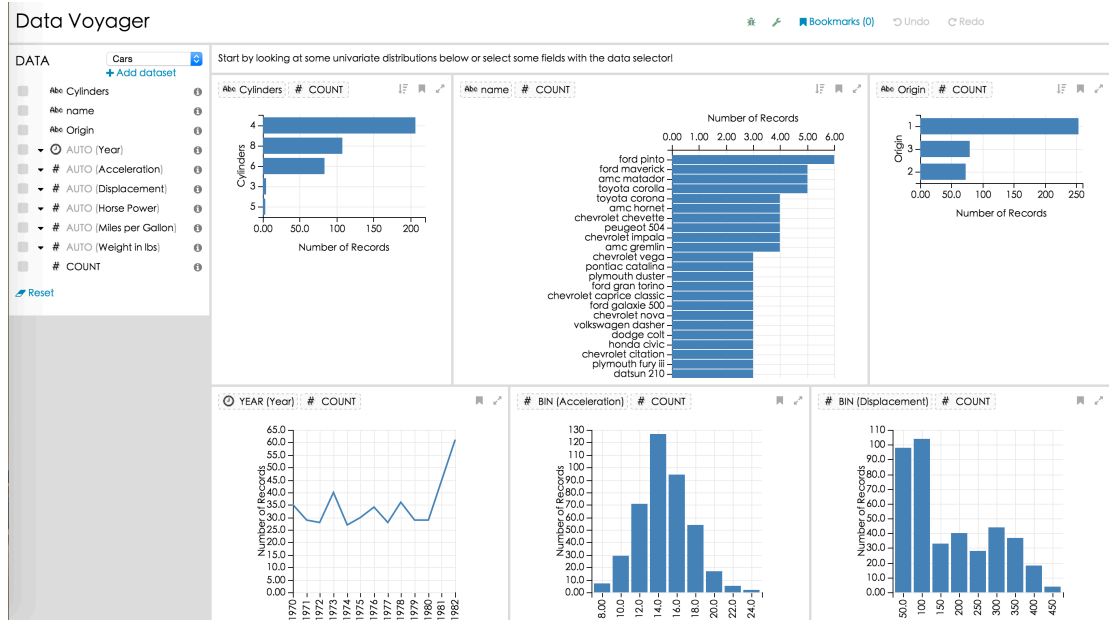
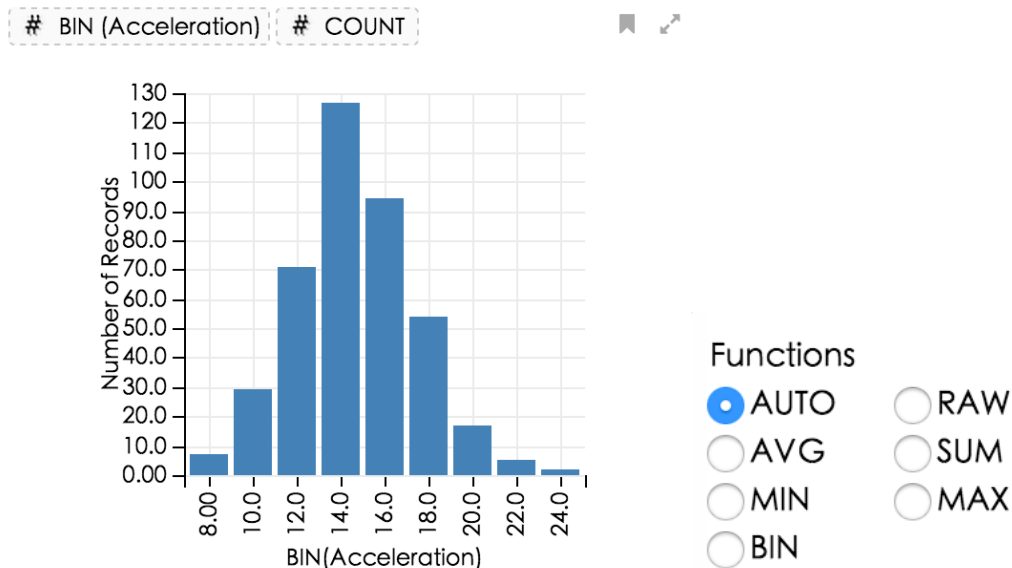


Data Voyager 调研报告

Data Voyagers 是网页端的数据可视化工具。对于一个多属性数据，Data Voyager 展示了数据在每个属性上的分布，并可以在用户交互下产生二维、三维的可视化展现。

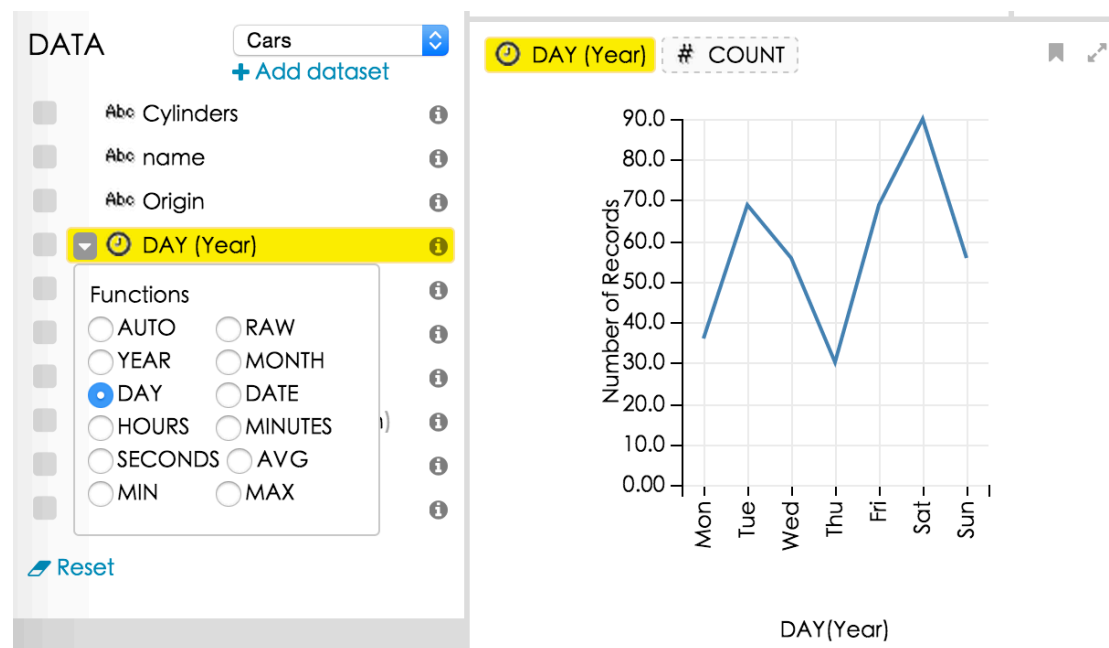


上图是打开 cars 数据集之后的初始化页面，左图和 tableau 类似，展示了数据属性及其类型（数值、类别、时序）。右图则展示了所有属性的基本分布，类别型和数值型为 bar chart，时序类型为 line chart。

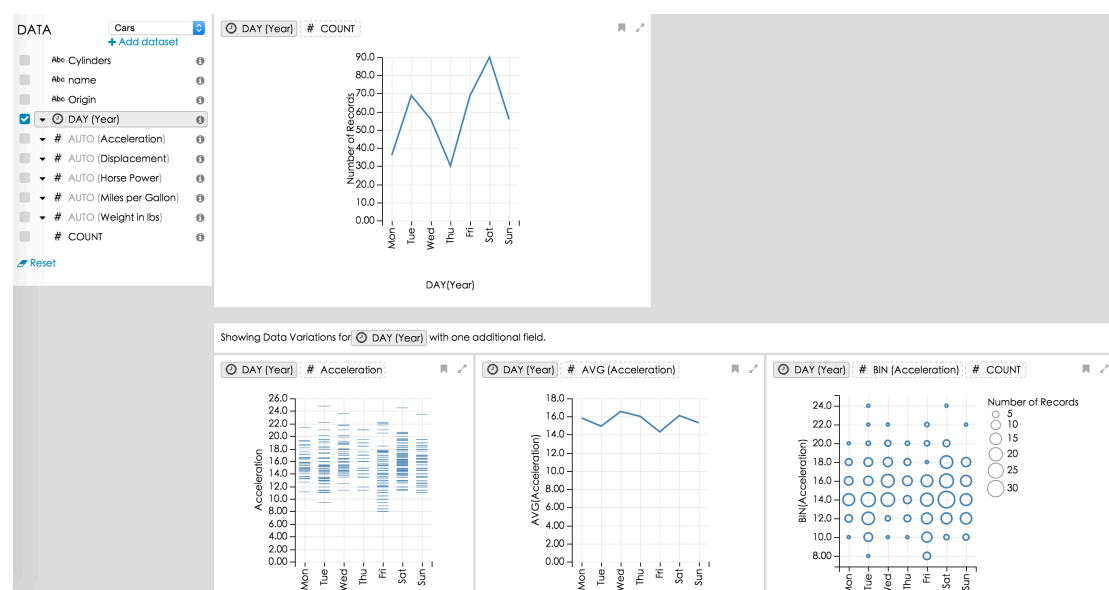


对于数值型属性，默认提供的是 binning 后的统计结果（如上图左），用户也可以调整为其它 aggregation 的结果（如上图右），但是这个 aggregation 结果是个

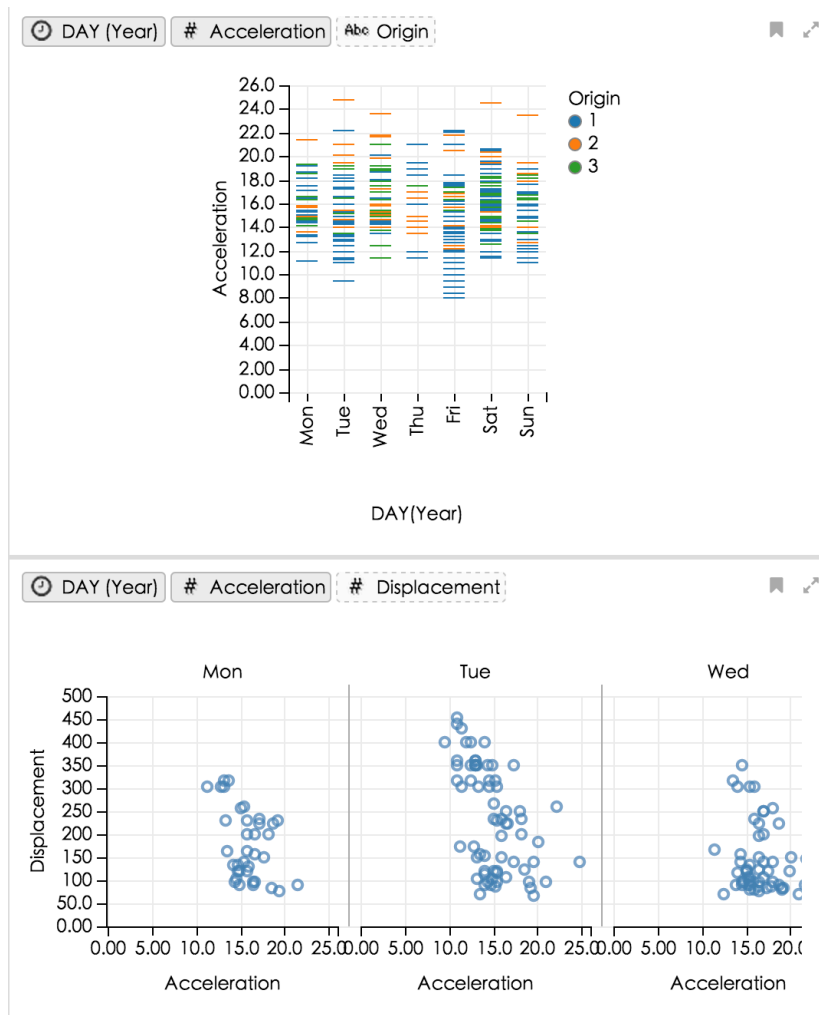
单一数值而不是序列，执行聚合操作后左图就会消失。



对于时序型属性，用户还可以选择丰富的按日、月、年、星期等聚合方法，对应的视图会产生相应变化。上图中把原来按年聚合的数据换成了按星期聚合。



当用户在左边的选择框勾选一个属性(如上图这里选择了 Year)，右边就会自动生成 Year 属性和其他属性的联合分布视图，每行是一个其他属性，从左到右分别是 Year 和其他属性的二维联合分布视图、Year 和其他属性的均值的分布视图和 Year+其他属性+count 的视图。上图的其他属性是数值型属性，如果是类别型属性就只有二维联合分布的 count 视图。



如果用户在左边勾选了 2 个属性，右边就会首先生成这两个属性的基本分布（就是上一段提到的那三个视图），然后下面是其他属性和这两个属性的 3 维视图，第三维度一般用颜色表达，或者把类别较少的维度按类似于散点矩阵的方法列举画出多个小视图的拼接。

再增加勾选的属性，也可以表达，但是在可视化上并不会对用户产生助力，反而增加了用户的阅读负担。

另外对于每个数据视图，用户可以执行重排序、视图最大化和收藏等操作。左边的选择框同时展现了每个维度的基本统计信息如最值、均值等。

对于数据的 scalability 问题，cars 数据集的数据量只有 406 个，网上提供的其他数据集也都是百数量级的。

以上就是我对 Data Voyager 的调研结果，在交互和细节设计上对 Data Scanner 有一定的借鉴作用，但是在数据量可扩展性上和分析上 Data Scanner 绝对更优。

